This is a working document prepared by the Energy Information Administration (EIA) in order to solicit advice and comment on statistical matter from the American Statistical Association Committee on Energy Statistics. This topic will be discussed at EIA's spring 2006, meeting with the Committee to be held April 6 and 7, 2006.

Appendix C.
Proposed Modifications to the EIA-914 Methodology

This section describes the proposed data estimation methodology used to estimate total production from respondent data. This will be a relatively qualitative presentation of this proposed methodology which focuses on the reduction of errors that result from assumptions in tested methodologies.

Gross Production Estimation for the Six Areas (Texas, Louisiana, Oklahoma, Wyoming, New Mexico, and Federal Gulf of Mexico)

A preliminary estimate of the final *Total Gross Production Rate* for each area is based on data provided by a cut-off sample of all operators for the data month. A cut-off sample was selected based on data for 2004.

Estimation

Gross Production Estimates for the Six Areas: A preliminary estimate of the final *Total Gross Production Rate* for each area (Texas, Louisiana, Oklahoma, Wyoming, New Mexico, and Federal Gulf of Mexico) is based on data provided by a cut-off sample of all operators for the data month. The preliminary total estimate is made each month by collecting gross production data from the sampled operators for the data month and adding to this an estimate of the gross production data from all operators *not* in the sample.

[1] $\hat{T}_t = S_t + \hat{N}_t$

This discussion will be focused on estimating the gross production each month,t, from all operators not in the sample, $\hat{N}_t$. A simple ratio model is given in equation [2] for any particular calibration year,c.

[2] $\hat{N}_t = (R_c) * (S_t)$

The value of R_c can assumed to be constant or variable over time. If assumed constant, it can be determined using variations of the classic Ratio Estimate Method for any area and time period for which the historical data are essentially complete. The ratio estimator, typically used for estimation with a cut-off sample, assumes that the sample coverage remains constant over time.

[3] $R_c = \frac{N_c}{S_c}$

As an example of this type of model, consider 2000 calibration year historical data:
Where

T_{00} = *Total Gross Production Rate in 2000=15,604 mmcf/day*,

S_{00} = *Sampled Operators Gross Production Rate* in *2000 = 13,658 mmcf/day*, and

N_{00} = *Not Sampled Operators Gross Production Rate* in 2000 = 1,945 mmcf/day.

Let

[4] $R_{00} = \frac{N_{00}}{S_{00}} = \frac{1{,}945}{13{,}658} = 0.1424$

For calibration year 2000, the model in equation [2] becomes

[5] $\hat{N}_t = 0.1424 * (S_{T,t})$

The estimate of $\hat{N}_t$, the Non-Sampled production, can be estimated from subsets of the total sampled production. The total Sample ratio model, along with a 1st Quartile model, the Upper 2 Quartile ratio model, and the Lower 2 Quartile ratio model are shown in Figure 3. The best performing constant ratio model was based on the Lower 2 Quartiles.

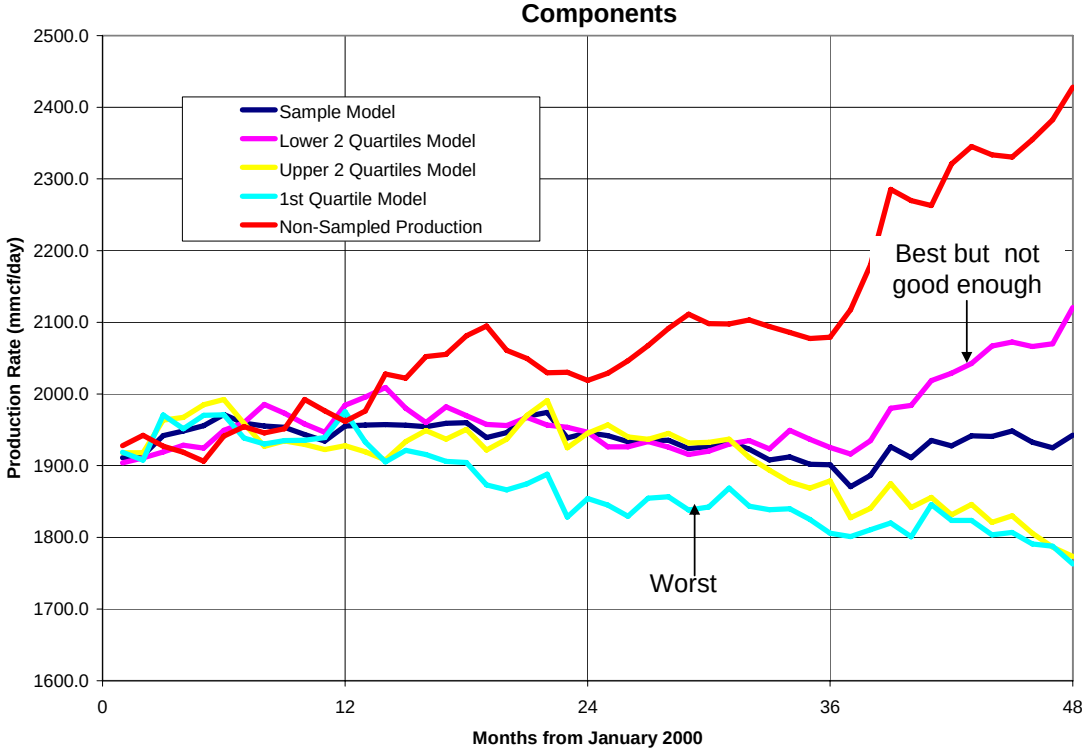
[6] $\hat{N}_t = 0.2921 * (S_L)$

The worst performing ratio model was based on the 1st Quartile of the sample production

[7] $\hat{N}_t = 0.5210 * (S_F)$

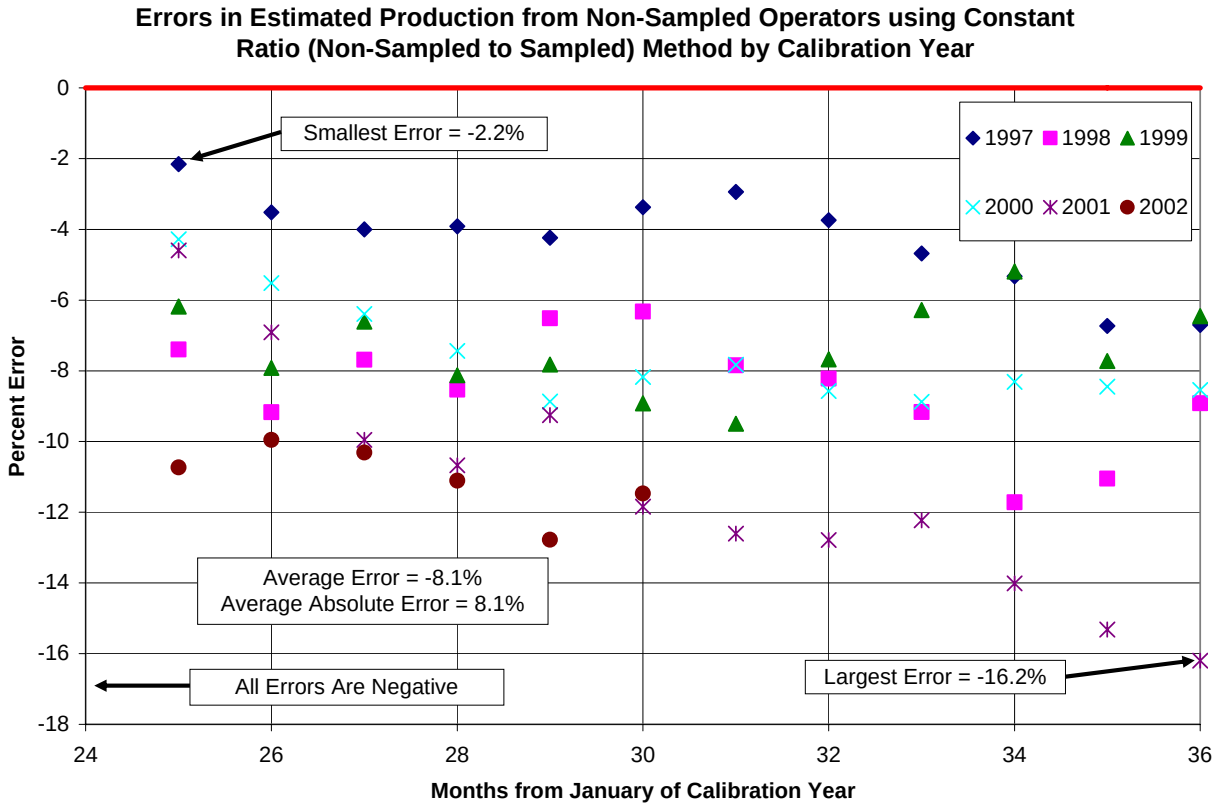
Similar results were obtained from the rest of the calibration years.

Figure 3. Constant Ratio Models of Non-Sampled Production by Sample Components

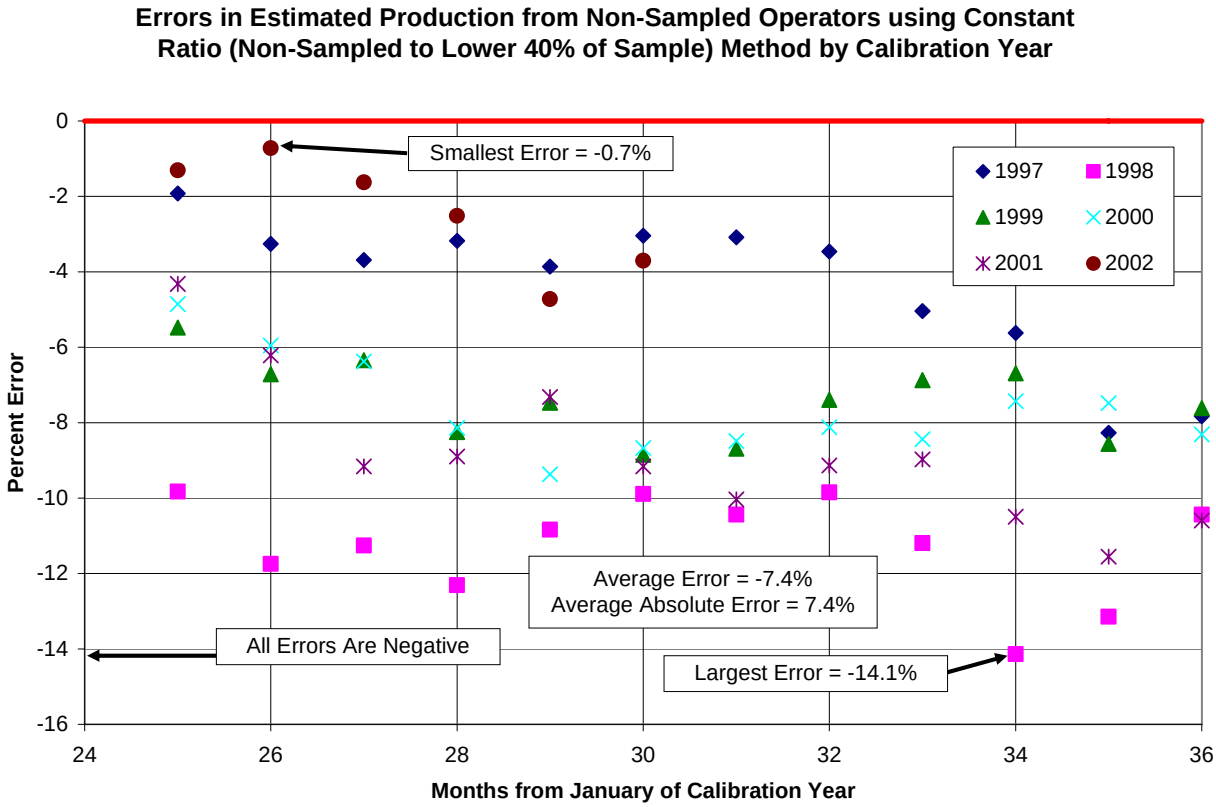


The preliminary total estimate will be made for each month in 2006 by collecting gross production data from the sampled operators for the data month, dividing by the number of days in a month to obtain an estimate for the gross production rate in billion cubic feet per day, and multiplying a subset of the sampled operators $S_{L,t}$ by a ratio.

The errors resulting from the various constant ratio methods were calculated.



The average absolute error and the largest error were somewhat lower when only the lowest e quartiles were used at 7.4 percent and minus 14.1 percent respectively.



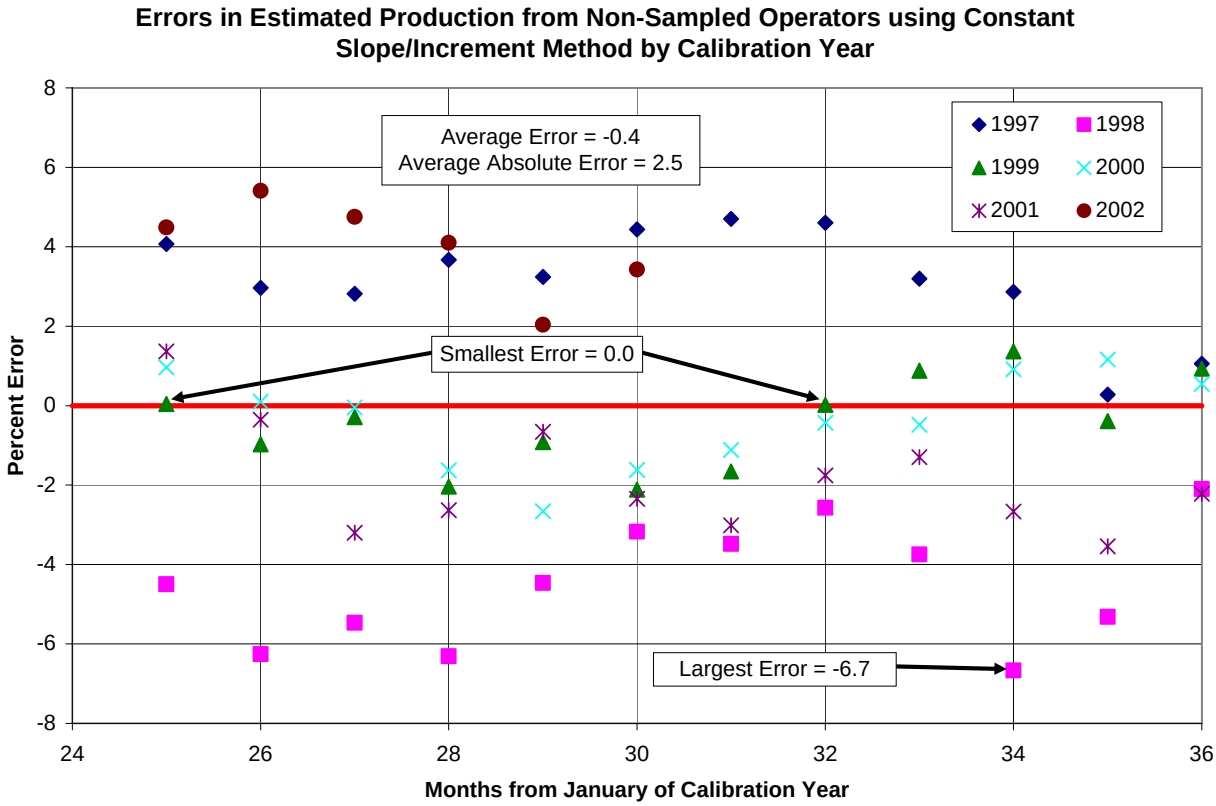
Models with variable ratios R_t were also tested.

[8] $\hat{N}_t = (\hat{R}_t) * (S_{L,t})$

These variable ratios had either constant or variable slopes. For the constant slope models,

[9] $R_t = (\hat{R} + \hat{a} * t) * (S_{L,t})$

where $\hat{R}$ and $\hat{a}$ are fit parameters. The errors associated with variable ratios were substantially smaller than those for constant ratio models. The average absolute error and the largest error were 2.5 percent and minus 6.7 percent respectively compared to the best constant ratio model errors of 7.4 percent and minus 14.1 percent respectively.



Somewhat better results were obtained using variable ratios that had variable slopes. Apparently the level of drilling for gas wells has a significant impact on the slopes. In equation **[10],** the D_t term depends on the level of drilling for natural gas at specific times.

[10] $R_t = (\hat{R} + \hat{a} * t + \hat{b} * [D_t] * t) * (S_{L,t})$

The average absolute error was 2.3 percent and the largest error in $\hat{N}_t$ was minus 5.6 percent. Remembering that $\hat{N}_t$ is less than 15 percent of the total production, the average absolute error in the estimated total production was less than 0.4 percent and the largest error in the six calibration years tested was less than one percent.

The sample selection and modeling will not lead to substantial errors. However, problems with survey data or basic calibration can lead to larger errors.

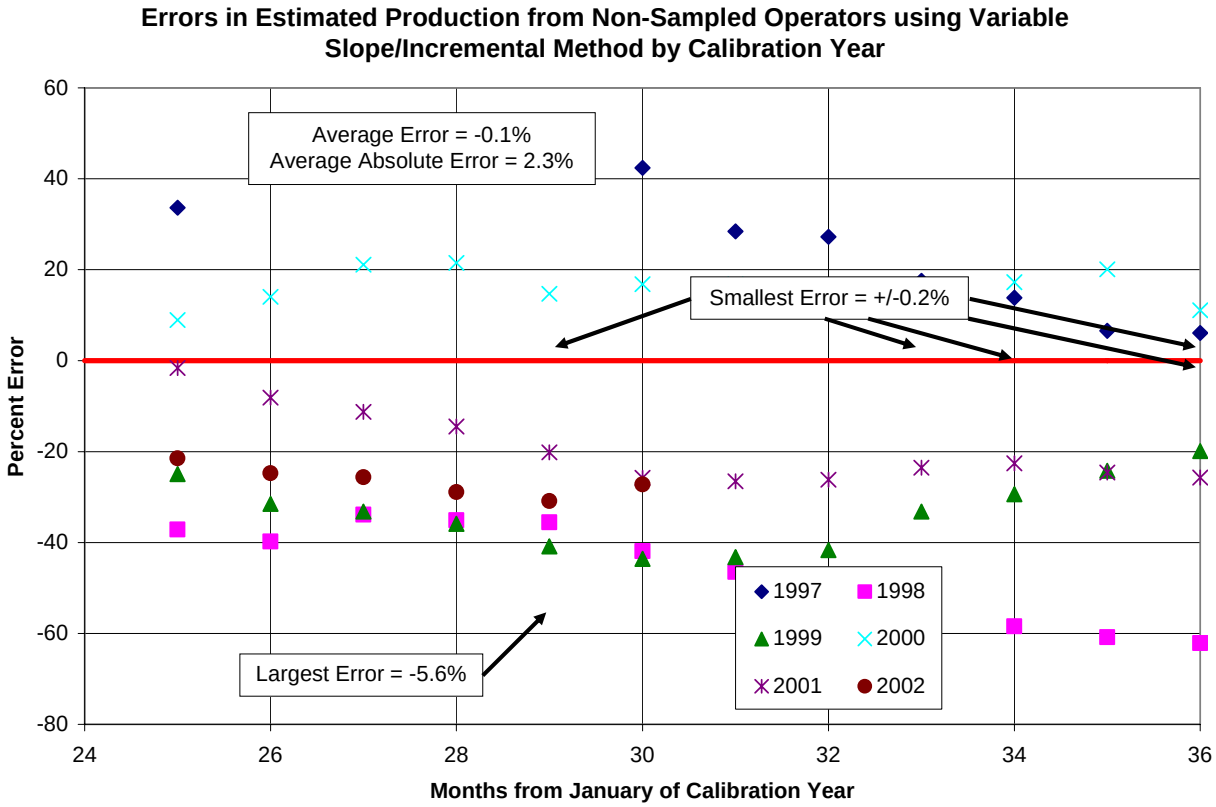

Errors in Estimated Production from Non-Sampled Operators using Variable Slope/Incremental Method by Calibration Year
Average Error = -0.1%
Average Absolute Error = 2.3%
Smallest Error = +/-0.2%
Largest Error = -5.6%
1997
1998
1999
2000
2001
2002
Percent Error
Months from January of Calibration Year
60
40
20
0
-20
-40
-60
-80
24
26
28
30
32
34
36

Figure 1